

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:44 ; Search time 170.72 Seconds  
(without alignments)  
15.909 Million cell updates/sec

Title: US-09-331-631a-8\_COPY\_80\_119  
Perfect score: 225  
Sequence: 1 PEDPQRREYECQECRQOEERQOPQCQQRCLKRFQEQEQ 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR66:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	221	98.2	588	1 FWCNAB	alpha-globulin B p
2	213	94.7	509	2 S08059	alpha-globulin typ
3	142	63.1	605	2 S06398	alpha-globulin typ
4	119	52.9	566	2 S22477	vicilin precursor
5	86	38.2	1905	2 T18267	multidrug resistanc
6	83	36.9	429	2 S29565	apolipoprotein A-I
7	82	36.4	1403	2 S24548	homeotic protein p
8	82	36.4	339	1 TWH02D	transcription init
9	82	36.4	401	2 A47141	apolipoprotein A-I
10	80	35.6	1457	2 T14577	protein kinase Yak
11	78	34.7	613	2 S27770	hypothetical prote
12	78	34.7	648	1 Q01150	protein kinase (EC
13	77	34.2	425	2 T18592	hypothetical prote
14	77	34.2	758	2 S54522	hypothetical prote
15	77	34.2	1761	2 T13675	hypothetical prote
16	76.5	34.0	388	2 T31887	hypothetical prote
17	76.5	34.0	388	2 T31888	hypothetical prote
18	76.5	34.0	467	1 A49377	glutamine-rich pro
19	76.5	34.0	738	2 S37876	involucrin - mouse
20	76	33.8	572	2 T29880	hypothetical prote
21	76	33.8	600	2 T18593	hypothetical prote
22	76	33.8	905	1 RGRV55	regulatory protein
23	76	33.6	1038	2 T02654	rep protein homolo
24	75.5	33.6	438	2 T31869	hypothetical prote
25	75.5	33.6	445	2 T31868	hypothetical prote
26	75	33.3	1023	2 T13068	CLOCK protein - fr
27	75	33.3	1027	2 T13071	CLOCK protein - fr
28	75	33.3	1154	2 S69206	regulator protein
29	75	33.3	1969	2 T08875	histidine kinase h

30	74.5	33.1	1390	2 T14004	trfa protein - sli
31	73.5	32.7	411	2 T29475	hypothetical prote
32	73	32.4	139	2 A26892	Mopa box protein -
33	73	32.4	646	2 D82493	conserved hypochet
34	73	32.4	1505	2 JC4851	hypoxia-inducible
35	72.5	32.2	292	2 B22364	alpha/beta-gliadin
36	72.5	32.2	307	2 S10015	alpha/beta-gliadin
37	72.5	32.2	319	2 C22364	alpha/beta-gliadin
38	72.5	32.2	326	2 D22364	alpha/beta-gliadin
39	72.5	32.2	1069	2 T00377	KIAA0642 protein -
40	72	32.0	538	2 S29521	casein kinase I ho
41	72	32.0	4957	2 T03455	ALR protein - huma
42	72	32.0	5262	2 T03454	ALR protein - huma
43	71	31.6	286	1 EEWTA	alpha/beta-gliadin
44	71	31.6	286	2 S07923	alpha/beta-gliadin
45	71	31.6	1365	2 S14871	suppressor two of

## ALIGNMENTS

## RESULT 1

FWCNAB  
N:Alternate names: seed precursor (clone C72) - upland cotton

A:Alternate names: seed storage protein; vicilin precursor



A:Accession: S24548  
A:Molecule type: DNA  
A:Residues: 1-1403 <DOE>  
A:Cross-references: EMBL:Z11743; NID:g8383; PID:g8384  
R:Matsuzaki, F.; Koizumi, K.; Hama, C.; Yoshioke, T.; Nabeshima, Y.  
Biochem. Biophys. Res. Commun. 182, 1326-1332, 1992  
A:Title: Cloning of the *Drosophila* prospero gene and its expression in ganglion mother cells  
A:Reference number: J01397; MUID:92171948  
A:Accession: J01397  
A:Molecule type: mRNA  
A:Residues: 1-417, 'Q', 419-1403 <MAT>  
A:Cross-references: GB:D10609; DDBJ:D90560; NID:g217345; PID:d1001939; PID:g217346  
R:Vaesslin, H.; Grell, E.; Wolff, E.; Bieri, E.; Jan, L.Y.; Jan, Y.N.  
Cell 67, 941-953, 1991  
A:Title: prospero is expressed in neuronal precursors and encodes a nuclear protein that is required for neurogenesis  
A:Reference number: M1089; MUID:92069760  
A:Accession: M1089  
A:Molecule type: mRNA  
A:Residues: 1-75, 'GQDAERAWPDPDAGPCRNEMPA', 99-119, 'NLAIQFHVQVAAAAATTTALLPIPG', 145-676, 'C  
A:Cross-references: GB:M61389; NID:g158183; PID:g158184  
C:Genetics:  
A:Gene: FlyBase:pros  
A:Cross-references: FlyBase:FBgn0004595  
C:Keywords: DNA binding; nucleus; transcription regulation  
F:717-734/Region: glutamine-rich  
F:991-998/Region: nuclear location signal  
F:1029-1049/Region: glutamine-rich

Query Match Similarity	36.9%	Score	83	DB	2	Length	1403
Best Local Similarity	54.8%	Pred	NO	0.23			
Matches	17	Conservative	7	Mismatches	7	Indels	0
QY	10	ECQCECRQDEERQDPCCQRCCLKRFEEQEQQ	40				
		:::					
db	716	EQQQQQQQQQQQQQQQQQQQEQQRFFEEQEQ	746				

RESULT 8  
TMW02D  
transcription initiation factor IID - human  
N:Alternate names: TATA-binding protein  
C:Species: Homo sapiens (man)  
C:Date: 20-Jul-1990 #sequence\_revision 19-May-1995 #text\_change 18-Feb-2000  
A:Accession: A34830; A34831; S10944; 160128  
R:Peterson, M.G.; Tanese, N.; Pugh, B.F.; Tjian, R.  
Science 248, 1625-1630, 1990  
A:Title: Functional domains and upstream activation properties of cloned human TATA bind  
A:Reference number: A34830; MUID:90302006  
A:Accession: A34830  
A:Molecule type: mRNA  
A:Residues: 1-339 <PEP>  
A:Cross-references: GB:M5654; NID:g339491; PIDN:AAA36731.1; PID:g339492  
R:Kao, C.C.; Lieberman, P.M.; Schmidt, M.C.; Zhou, Q.; Pei, R.; Berk, A.J.  
Science 248, 1645-1649, 1990  
A:Title: Cloning of a transcriptionally active human TATA binding factor.  
A:Reference number: A34831; MUID:90302010  
A:Accession: A34831  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-17, 'N', '19-186, 'R', 188-339 <KAO>  
R:Horifumi, A.; Sinn, E.; Yamamoto, T.; Wang, J.; Roy, A.; Horikoshi, M.; Roeder, R.G.  
Nature 346, 387-390, 1990  
A:Title: Highly conserved core domain and unique N terminus with presumptive regulatory  
A:Reference number: S10944; MUID:90326195  
A:Accession: S10944  
A:Molecule type: mRNA  
A:Residues: 1-91,96-339 <HOF>  
A:Cross-references: EMBL:X54993; NID:g37065; PIDN:CAA38736.1; PID:g37066  
R:Kao, C.; Lieberman, P.; Schmidt, M.; Zhou, Q.; Pei, R.; Berk, A.J.  
Science 248, 1626, 1990  
A:Title: Cloning of the human TATA binding factor: Expression of a transcriptionally act  
A:Reference number: 160128

A:Accession:160128  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-186, 'R', 188-299, 'MIKPR', 300-339 <RES>  
A:Cross-references: GB:M34960; NID:g339493; PID:g339494  
C:Genetics:  
C:Gene: GDB:TBP; GTF2D1  
A:Cross-references: GDB:138768; OMIM:600075  
A:Map position: 6927-6927  
C:Superfamily: human transcription initiation factor IID  
C:Keywords: alternative splicing; DNA binding; nucleus; transcription initiation  
C:55-95/Region: glutamine-rich

		36.4%;	Score	82;	DB	1:	Length	339;	
Query Match		41.0%;	Pred.	No.	0.09;				
Best Local Similarity									
Matches	16;	Conservative	14;	Mismatches	9;	Indels	0;	Gaps	0;
OY	2	EDPQRRECCOECNOQEBRODPQCOCRLKRFBDQBDOO	40						
	:	: : :	:	:	:	:	:	:	:
db	53	EEOGROROOOONNNNNNNDNNNNNNNNNNNNNNNNNNNN	91						

```

RESULT      9
A47141
apolipoprotein A-IV I isoform - baboon (fragment)
C:Species: Papio sp. (baboon)
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 12-Apr-1995
C:Accession: A47141
R:Hixson, J.E.; Kammmerer, G.M.; Mott, G.E.; Britten, M.L.; Birnbaum, S.; Powers, P.K.
J. Biol. Chem. 268, 15667-15673, 1993
A:Title: Baboon apolipoprotein A-IV. Identification of Lys76-->Glu that distinguishes
A:Reference number: A47141; MUID:93340170
A:Accession: A47141
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-401 <HTX>
A:Experimental source: intestine
A:Note: Sequence extracted from NCBI backbon (NCBIN:136009, NCBIIP:136010)
;Superfamily: apolipoprotein A-I

```

[illegible]

```

RESULT 10
T14577
protein kinase Yaka (EC 2.7.1.-) - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
A:Accession: T14577
R:Kuspa, A.; Lu, S.; Souza, G.M.
submitted to the EMBL Data Library, January 1998
A:Description: Yaka, a protein kinase required for the growth to development transmitti
A:Reference number: Z18146
A:Accession: T14577
A>Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1457 <KUS>
A:Cross-References: EMBL:AF045453; NID:g2854116; PTD:g2854117; PIDN:AAC02554.1
C:Genetics:
A:Gene: yaka
C:Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protei

```

Matches 15; Conservative 14; Mismatches 10; Indels 0; Gaps 0;  
Query 2 EDPORRYEECCQOECCROOEERQOPQCCORCLKRFEEQEQ 40  
Db 598 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 636

## RESULT 11

S27770  
hypothetical protein 1 - African malaria mosquito (fragment)  
C:Species: Anopheles gambiae (African malaria mosquito)  
C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 09-Sep-1997  
C:Accession: S27770  
R:Beasansy, N.J.; Paskewitz, S.M.; Mills-Hamm, D.M.; Collins, F.H.  
submitted to the EMBL Data Library, June 1992  
A:Description: Distinct families of site-specific retroposons occupy identical positions  
A:Reference number: S27770  
A:Accession: S27770  
A:Molecule type: DNA  
A:Residues: 1-613 <BES>  
A:Cross-references: EMBL:M93690; NID:g159615; PID:g159616

Query Match 34.7%; Score 78; DB 2; Length 613;  
Best Local Similarity 47.2%; Pred. No. 0.39;  
Matches 17; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

Query 5 QRRYEECCQOECCROOEERQOPQCCORCLKRFEEQEQ 40  
Db 230 QQREQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 265

## RESULT 12

JQ1150  
protein kinase (EC 2.7.1.37) CAMP-dependent, catalytic chain - slime mold (Dictyostelium  
C:Species: Dictyostelium discoideum  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 24-Apr-1998  
C:Accession: JQ1150  
R:Berri, E.; Anjard, C.; Scholder, J.C.; Raymond, C.D.  
Gene 102, 57-65, 1991  
A:title: Isolation of two genes encoding putative protein kinases regulated during Dicty  
A:Reference number: JQ1150, MIMD:91323730  
A:Accession: JQ1150  
A:Molecule type: DNA  
A:Residues: 1-648 <BUE>  
A:Cross-references: GB:M38703  
C:Genetics:  
A:Gene: PK2  
A:Introns: 578/3  
C:Complex: heterodimer with regulatory chain; active catalytic chain is released when c

A:Description: catalyzes the formation of peptidyl-serine-phosphate or peptidyl-threonin  
A>Note: Important for cell type differentiation and fruiting body morphogenesis  
C:Superfamily: Dictyostelium CAMP-dependent protein kinase catalytic chain; protein kin  
C:Keywords: ATP; magnesium; phosphoprotein; phosphotransferase; serine/threonine-specifi  
F:126-223/Region: glutamine-rich  
F:297-312/Region: glutamine-rich  
F:334-590/Domain: protein kinase homology <KIN>  
F:342-350/Region: protein kinase ATP-binding motif  
F:365,384,459,461/Active site: Lys, Glu, Asp, Lys #status predicted  
F:464,468/Binding site: magnesium (Asn, Asp) #status predicted

Query Match 34.7%; Score 78; DB 1; Length 648;  
Best Local Similarity 37.5%; Pred. No. 0.4;  
Matches 15; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Query 1 PEDQRRYEECCQOECCROOEERQOPQCCORCLKRFEEQEQ 40  
Db 143 PQQQQPQQQQPQQQQPQQQQQQQQQQQQQQQQQQQQQQQQ 182

## RESULT 13

T18592  
hypothetical protein AC3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T18592  
R:McMurray, A.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: T18592  
A:Accession: T18592  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-425 <WIL>  
A:Cross-references: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3  
A:Experimental source: clone AC3  
C:Genetics:  
A:Gene: CESP:AC3.3  
A:Map position: 5  
A:Introns: 18/3  
C:Superfamily: gliadin

Query Match 34.2%; Score 77; DB 2; Length 425;  
Best Local Similarity 38.9%; Pred. No. 0.36;  
Matches 14; Conservative 10; Mismatches 6; Indels 6; Gaps 1;

Query 4 PQRRYEECCQOECCROO-----EERQOPQCCORCLK 33  
Db 148 PVQCVPCQCCQCCQCCQVCTPPIQQCPQCCQCCQCCVQQ 183

## RESULT 14

S54522  
hypothetical protein YMR164c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YMR520.13c  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Oct-1999  
C:Accession: S54522; S54609  
R:Hunt, S.; Bowman, S.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S54510  
A:Accession: S54522  
A:Molecule type: DNA  
A:Residues: 1-758 <HUN>  
A:Cross-references: GB:Z49705; EMBL:Z49700; NID:g825556; PIDN:CAA89800.1; PID:g825569  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: SGD:MSS11  
A:Cross-references: SGD:S0004774; MIPS:YMR164c  
A:Map position: 13R

Query Match 34.2%; Score 77; DB 2; Length 758;  
Best Local Similarity 40.5%; Pred. No. 0.58;  
Matches 15; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Query 4 PQRRYEECCQOECCROOEERQOPQCCORCLKRFEEQEQ 40  
Db 293 PQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 329

## RESULT 15

T13675  
hypothetical protein EG0002.3 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jun-2000  
C:Accession: T13675  
R:Boisakov, V.; Borkova, D.; Minana, B.; Kafatos, F.  
submitted to the EMBL Data Library, September 1998  
A:Description: Sequencing the distal x chromosome of Drosophila melanogaster.  
A:Reference number: T13675  
A:Accession: T13675  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA



